Google Colab hoặc dùng Jupiter (trong Anaconda có)

https://www.youtube.com/watch?v=UaellPmi8CU&t=289s

# GIỚI THIỆU

Hiện nay bóng đá là một môn thể thao được yêu thích trên thế giới. Việc giao lưu thể thao mang lại sức khỏe cho cầu thủ và đem lại nhiều lợi ích: Thắt chặt mối quan hệ giữa các quốc gia, khẳng định vị thế tinh thần của một quốc gia, các liên đoàn, đội tuyển. Việc chuyển nhượng cầu thủ nhằm nhu cầu cho đội bóng hoặc câu lạc bộ trở nên mạnh hơn là một điều thiết yếu. Sự cạnh để đạt được các vị trí cầu thủ giữa các đội bóng là một điều quan trong dùng để phân tích. Môn học mạng xã hội có thể phân tích mạng chuyển nhượng cầu thủ bằng các độ đo (dựa trên đường đi, phổ, khoảng cách,...), khai thác dữ liệu trên mạng này (gom cụm).

# XÁC ĐỊNH BÀI TOÁN

Input: Tập dữ liệu ban đầu trên nguồn dữ liệu Kaggle đã qua tiền xử lý dữ liệu

Ouput: Đưa ra độ đo, đưa ra cộng đồng phục vụ cho việc phân tích mạng xã hội “Top 250 Football transfers from 2000 to 2018”

1. III. DỮ LIỆU

## Giới thiệu nguồn dữ liệu

Link dataset: https://www.kaggle.com/vardan95ghazaryan/top-250-football-transfers-from-2000-to-2018

Dữ liệu Top 250 Football transfers from 2000 to 2018 được cung cấp trên nền tảng Kaggle chứa dữ liệu việc chuyển nhượng cầu thủ từ năm 2000 đến 2018. Dữ liệu gồm 3104 dòng với 10 thuộc tính bao gồm:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *STT* | *Thuộc tính* | *Kiểu dữ liệu* | *Mô tả* | *Các giá trị* |
| *1* | *Name of a football player* | *char* | *Tên cầu thủ chơi bóng* |  |
| *2* | *Position* | *char* | *Vị trí chơi* | *Centre-Forward*  *Centre-Back*  *Other* |
| *3* | *Age* | *int* | *Tuổi* |  |
| *4* | *Team\_from* | *Char* | *Tên đội bán* |  |
| *5* | *League\_from* | *Char* | *Liên đoàn bán* |  |
| *6* | *Team\_to* | *Char* | *Tên đội mua* |  |
| *7* | *League\_to* | *Char* | *Liên đoàn mua* |  |
| *8* | *Season* | *Char* | *Mùa giải* |  |
| *9* | *Market Value* | *Float* | *Giá thị trường* |  |
| *10* | *Transfer\_fee* | *Float* | *Phí chuyển nhượng* |  |

* 1. Xử lý và phân tích dữ liệu

### Làm sạch dữ liệu

Bộ dữ liệu nhìn chung có vài thuộc tính bị thiếu dữ liệu,cần loại bỏ khỏi dataset.

**

*Hình 1 đọc dữ liệu từ file csv và loại bỏ dữ liệu trùng và rỗng*

Dữ liệu sẽ được đọc từ file csv đưa vào dataframe. Sau đó xoá bỏ dữ liệu trùng lặp và rỗng đi. Kết quả nhận được là một bộ dữ liệu 245 dòng và 2 cột

* + 1. Chuyển đổi dataframe thành đồ thị
* *Đồ thị 2 phía*

*Node: là vị trí các cầu thủ trên sân (Position) và đội tuyển mua cầu thủ đó(Team\_to), Edge: mối quan hệ giữa việc đội tuyển mua một vị trí trong đội hình bóng đá.*

*Ảnh có chứa văn bản

Mô tả được tạo tự động*

*Hình 2 đưa dữ liệu từ dataframe vào đồ thị vô hướng*

*Có thể thấy có 13 vị trí trên sân*

*Số đội tuyển là 108*

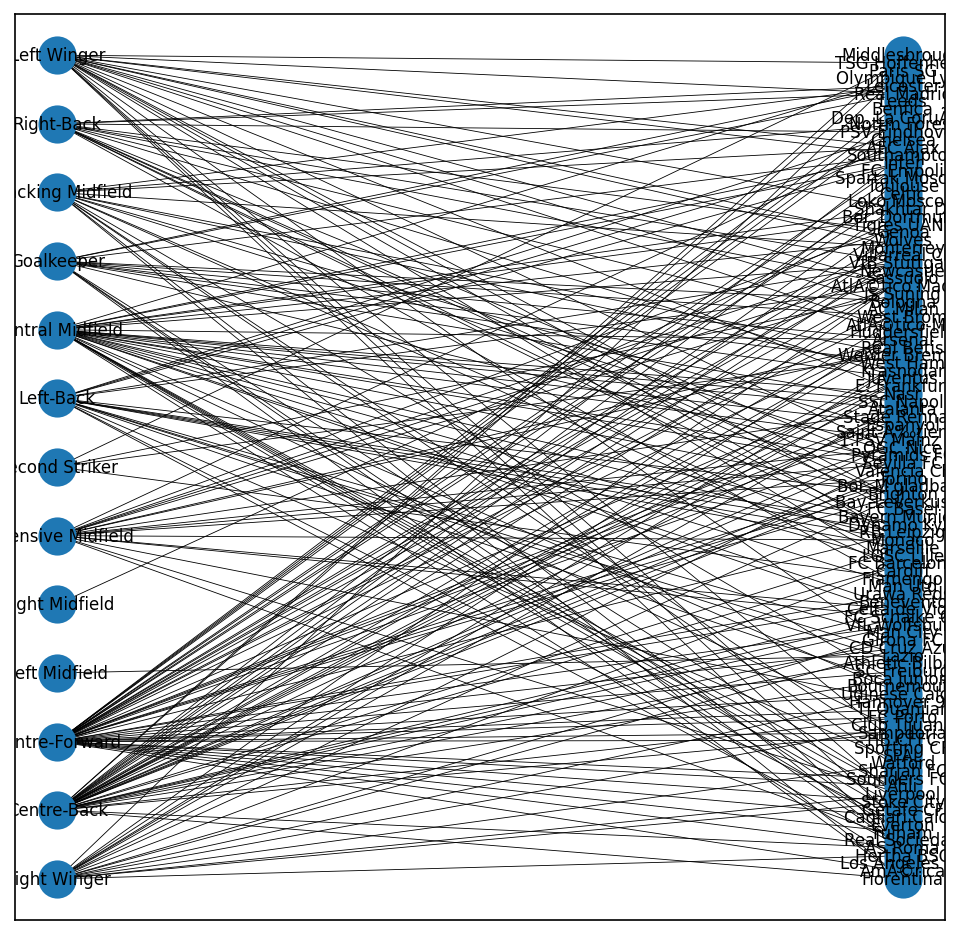
*Số cạnh là 245*

*- Code hiển thị đồ thị 2 phía*

*Ảnh có chứa văn bản

Mô tả được tạo tự động*

*Hình 3 code hiển thị đồ thị 2 phía*

**

*Hình 4 Đồ thị 2 phía*

*Nhìn vào đồ thị có thể thầy một đội vòng mua nhiều vị trí trong kỳ chuyển nhượng cầu thủ và một vị trí cũng sẽ được nhiều đội bóng để ý*

*- Đồ thị 1 phía*

*Node: là đội tuyển có mua các cầu thủ (Team\_to)*

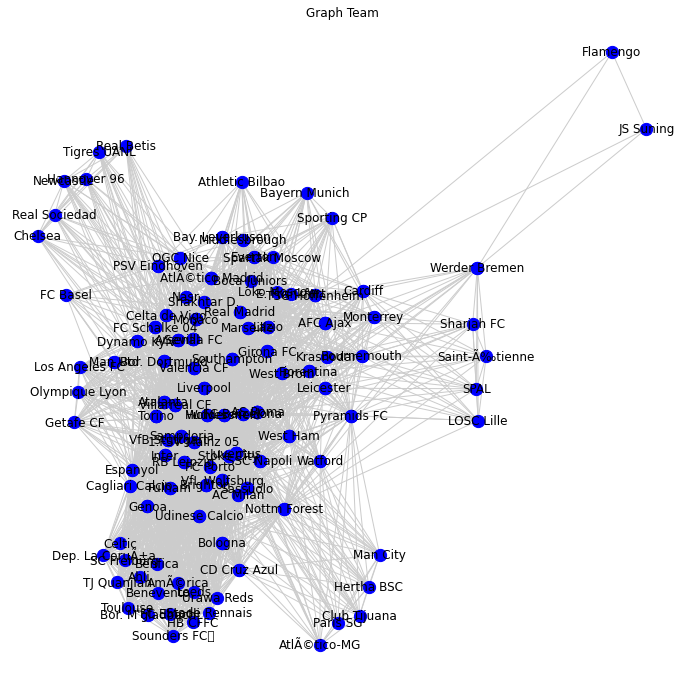
*Edge: Hai đội tuyển mua cùng một vị trí sẽ được nối với nhau để tạo thành cạnh, ý nghĩa nói lên sự cạnh tranh giữa hai đội tuyển trên thị trường chuyển nhượng khi mua cùng một vị trí*

*Weight : Trọng số là số vị trí trùng nhau mà hai đội tuyển đều mua*

*Ảnh có chứa văn bản

Mô tả được tạo tự động*

*Hình 5 Code hiển thị đồ thị 1 phía*

**

*Hình 6 Đồ thị 1 phía*

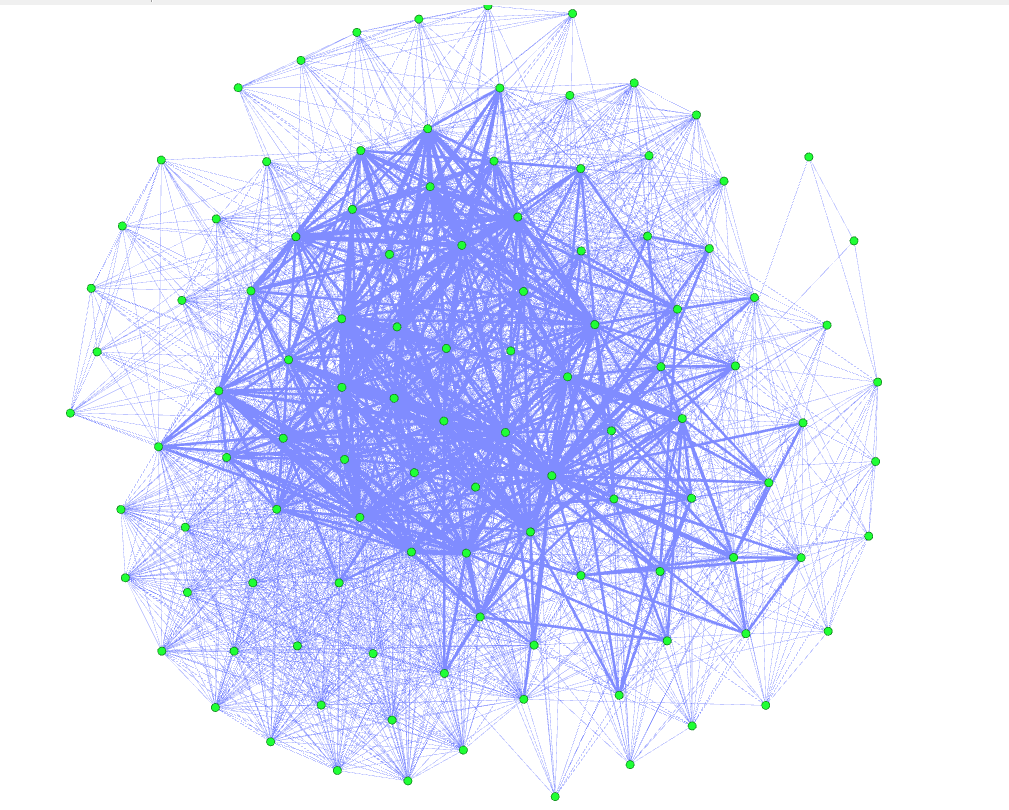
* *Xuất dữ liệu đồ thị 1 phía ra file csv để thực hiện trên Gephi*

*Ảnh có chứa văn bản

Mô tả được tạo tự động*

*Hình 7 Xuất dữ liệu đồ thị 1 phía ra file csv*

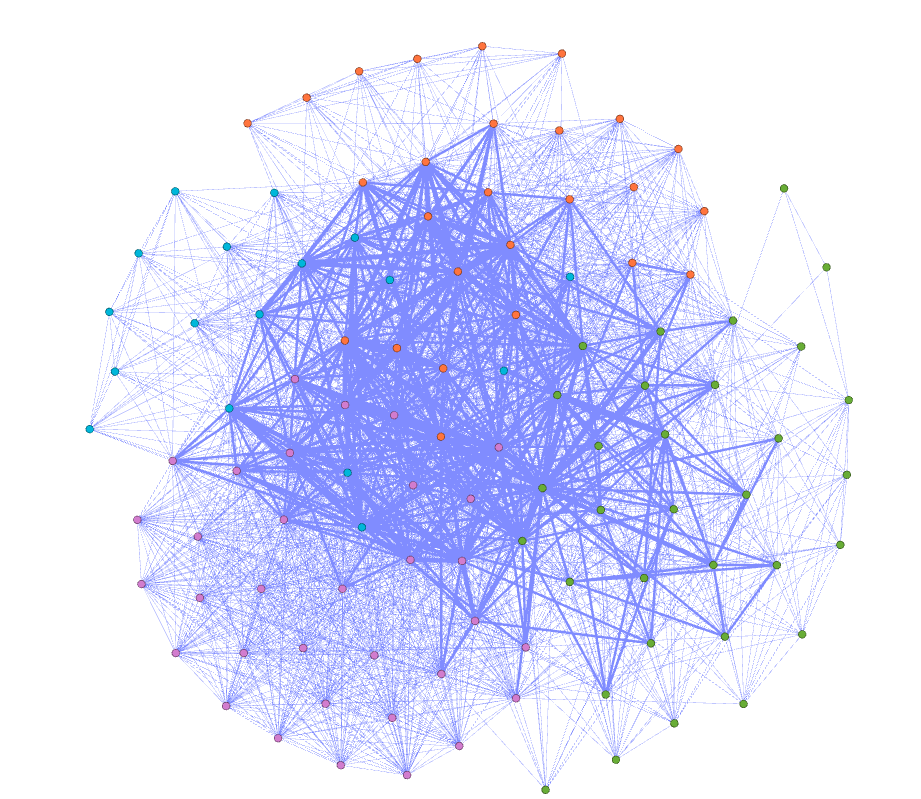
* *Đồ Thị 1 phía trên Gephi*

**

*Hình 8 đồ thị một phía trên Gephi*

# THUẬT TOÁN PHÁT HIỆN CỘNG ĐỒNG

## IV.1 Thực thi thuật toán Louvain bằng Gephi

**

*Hình 9 Đồ thị bằng Louvain*

*Dễ thấy thuật toán rút ra được 4 cụm*

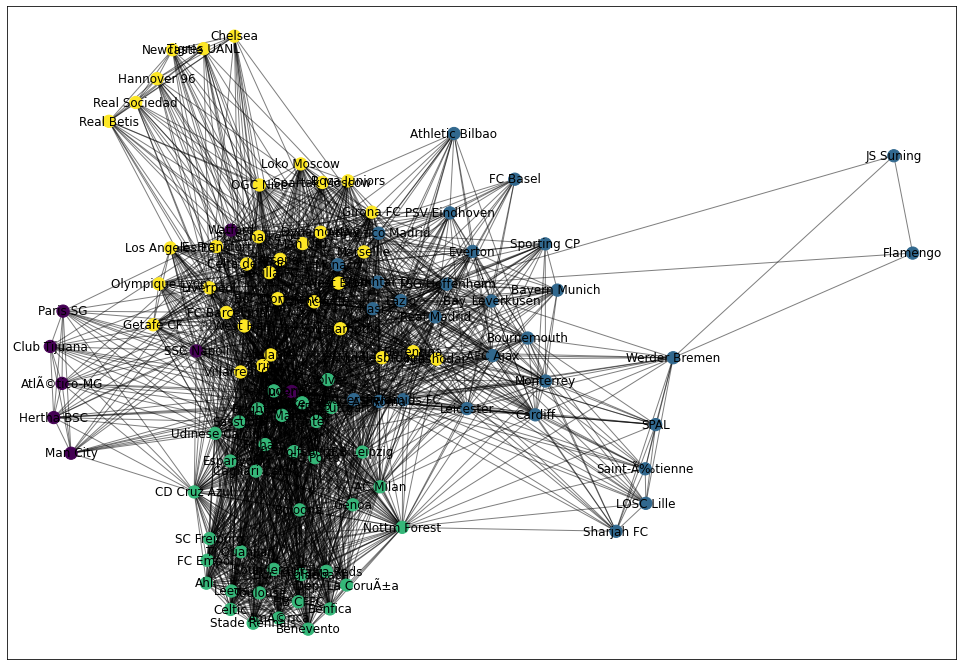
### IV.2 Thực thi bằng Python

import matplotlib.cm as cm  
import matplotlib

*Ảnh có chứa văn bản

Mô tả được tạo tự động*

*Hình 10 Code chạy thuật toán Louvain*

**

*Hình 11 Đồ thị phân cụm sử dụng Louvain*

*Ảnh có chứa văn bản

Mô tả được tạo tự động*

*Hình 12 danh sách các cụm*

*Số cụm: 4*

*Cụm thứ 0: Paris SG,Juventus,Man City,SSC Napoli,Watford,Hertha BSC,AtlÃ©tico-MG,Club Tijuana*

*Ý nghĩa cụm thứ 0: …….(SV bổ sung)*

*Cụm thứ 1: AtlÃ©tico Madrid, Real Madrid, Everton, Monaco, Leicester, AS Roma, Athletic Bilbao, Huddersfield, Bay. Leverkusen, AFC Ajax, Lazio, Monterrey, Werder Bremen, Bournemouth, Cardiff, Flamengo, Bayern Munich, Pyramids FC, Nasr, TSG Hoffenheim, Saint-Ã‰tienne, Shakhtar D., Sporting CP, Sharjah FC, JS Suning, PSV Eindhoven, LOSC Lille, SPAL, FC Basel*

*Ý nghĩa cụm thứ 1: …..*

*Cụm thứ 2: Inter, Fulham, TJ Quanjian, Bor. M'gladbach, AC Milan, Udinese Calcio, Brighton, Sampdoria, RB Leipzig, Nottm Forest, Wolves, VfL Wolfsburg, Stoke City, Genoa, Celtic, Ahli, Cagliari Calcio, Bologna, Sassuolo, Espanyol, VfB Stuttgart, Stade Rennais, FC Empoli, AmÃ©rica, FC Porto, 1.FSV Mainz 05, Leeds, Benfica, Sounders FC , Dep. La CoruÃ±a, HB CFFC, CD Cruz Azul, Toulouse, Urawa Reds, SC Freiburg, Benevento*

*Ý nghĩa cụm thứ 2: …….(SV bổ sung)*

*Cụm thứ 3: Liverpool, Man Utd, Chelsea, FC Barcelona, West Ham, Arsenal, Bor. Dortmund, Valencia CF, Southampton, Real Betis, Villarreal CF, Marseille, Sevilla FC, Torino, Real Sociedad, FC Schalke 04, OGC Nice, Fiorentina, Tigres UANL, Olympique Lyon, Spartak Moscow, Middlesbrough, Krasnodar, Celta de Vigo, West Brom, Newcastle, Atalanta, Hannover 96, Los Angeles FC, Boca Juniors, Loko Moscow, Getafe CF, Dynamo Kyiv, Girona FC, E. Frankfurt*

*Ý nghĩa cụm thứ 3: …….(SV bổ sung)*

# XẾP HẠNG CÁC ĐỘI BÓNG

## a) Thuật toán Pagerank (Thực thi bằng Gephi)

*Ảnh có chứa bàn

Mô tả được tạo tự động*

*Hình 13 Top đội bóng xếp hạng theo pagerank dung Gephi*

*Kết quả có không có sự khác biệt giữa code python và Gephi. Hai kết quả trùng khớp từ các thông số đến thứ hạng*

## b) Thuật toán closeness\_centrality (Thực thi bằng Gephi)

*Kết quả của closeness centrality thể hiển top độ gần các node đến tất cả các node trong mạng. Nó cũng đồng nghĩa với việc top các đội bóng phải cạnh tranh khốc liệt, mức độ cạnh tranh cao nhất*

*Ảnh có chứa bàn

Mô tả được tạo tự động*

*Hình 14 Top đội bóng xếp hạng theo closeness centrality*

***(Phần code python các bạn tự làm)***

1. TÀI LIỆU THAM KHẢO

*[1] https://en.wikipedia.org/wiki/Betweenness\_centrality*

*[2] https://en.wikipedia.org/wiki/Louvain\_method*